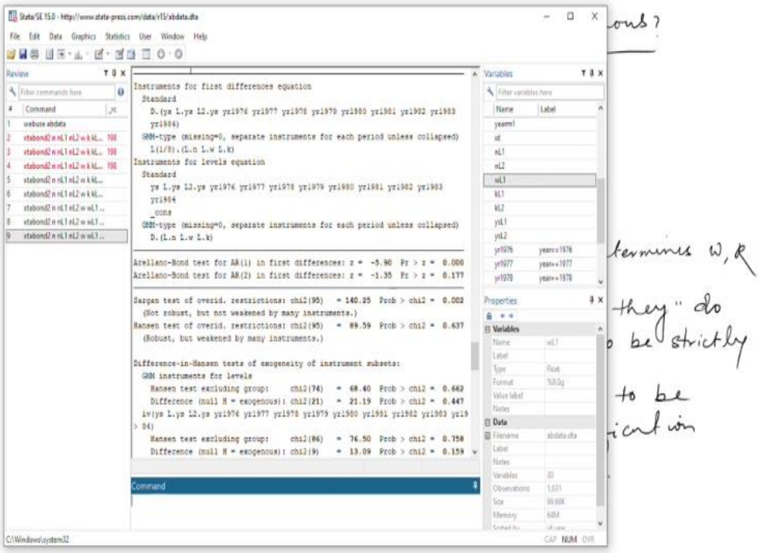
Applied Econometrics
Prof. Sabuj Kumar Mandal
Department of Humanities and Social Sciences
Indian Institute of Technology – Madras

Lecture - 51
Dynamic Panel Data Model – Part 14

(Refer Slide Time: 00:15)



This is what we have understood. Just by doing the specification properly, we have gained a lot. What will happen if we do system GMM? **(Video Starts: 00:32)** System GMM means I will estimate *nL1* which was lying earlier within the interval *xtabond2 n nL1 nL2 w wL1 L(0/2). (k ys) yr*,gmmstyle(L.(n w k)) ivstyle(L(0/2).ys yr*) robust small* now coefficient it is going to 1.06. So probably what I have to do is when I am doing this system GMM then I need to estimate instead of here what I am doing. I am doing a one-step system GMM.

So, the one-step system GMM, it has a problem, what is the problem? I will discuss that later. So, for the time being, I will estimate a two-step system GMM model. So here I will simply put two steps *xtabond2 n nL1 nL2 w wL1 L(0/2). (k ys) yr*, gmmstyle(L.(n w k)) ivstyle(L(0/2).ys yr*) twostep smal lnocons* and I do not require *robust*, this is a system GMM that we are estimating we need to put that no cons command also because no cons will remove the constant from the x and z and then if we do so it is just lying 1.08, it is still outside the bound.

That means with this example, what we are trying to say is that the specification actually matters a lot. Here I am saying I am getting 1.08 again which is lying outside the interval, but when we have estimated a difference GMM, one step difference GMM within that specification it is lying within the interval. What I am trying to convey here is whatever model we estimate whether we estimate a difference GMM.

Whether we estimate a system GMM, whether what variable we specify as endogenous, what variable we specify as exogenous those things actually matter a lot. So, one thing, what you have to do is the moment we get a model we first get its OLS estimation, then we will get its fixed effect estimation to get the limit, upper limit and lower limit. Once we get the limit, then we will estimate the model, either system or difference, either one step or two steps.

After that, we will see whether the estimates are lying within the interval or not. Until and unless our estimates are lying within the interval, we need to keep on changing our model specification, here we have changed two variables, as earlier they are used as exogenous, and now they are endogenous. Now how do you know that those variables should be exogenous or indigenous? Probably the technique what we have learned earlier.

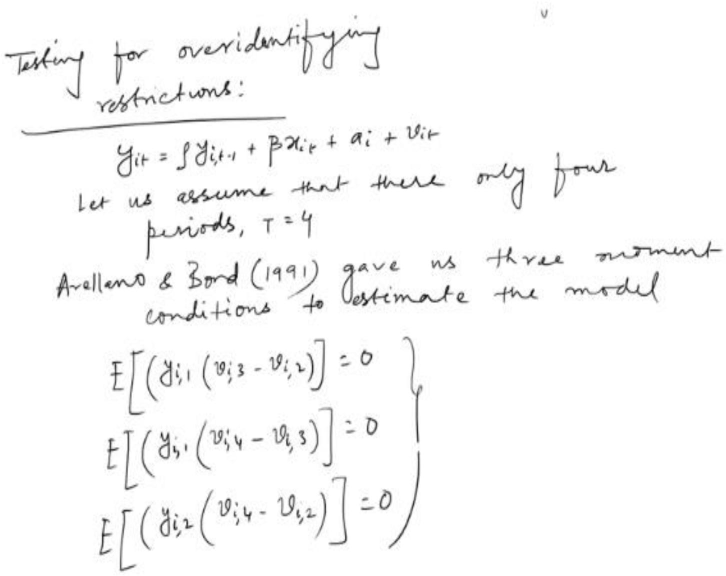
We have to run a simple panel data model fixed effect panel data model and then some tests of endogeneity like the Wu-Hausman test or Hausman test of endogeneity. We should run some kind of endogeneity test to check whether those variables are actually endogenous or not. Depending on those results, we will put our final specification in the context of this dynamic panel and then until and unless our estimates are lying within the interval we need to play around with our specification.

Here we have played only with changing exogenous and endogenous variables, probably it also depends on how many lags you are using as your instrument. Look at here, here the number of instruments are 90. When we are using this model number of instruments in system GMM is 112, so it is very difficult to ensure overidentifying restrictions for all these 112 instruments. So, we will learn how to restrict the number of instruments as well later on.

For the time being what I should also convey to you, so far we are always talking about overidentifying restrictions, whether all the instruments are satisfying overidentifying restrictions or not, before we talk about those that means the second post-estimation checkup even though they are satisfied here, look at here overidentifying restriction, Hausman test of overidentifying restriction is 0.637. So that means what is the null hypothesis?

The null hypothesis is that overidentifying restrictions are valid, so we are unable to reject the null, but Sargan rejects the null hypothesis. **(Video Ends: 06:25)** Before we talk about this, let us now take a little bit of theoretical concept which will help you to understand what exactly is the overidentification we are talking about in the context of a GMM estimation.

(Refer Slide Time: 06:59)



So here let us talk about now testing for overidentifying restrictions that we are saying testing for overidentifying restrictions. This is our model $y_{it} = \rho y_{i,t-1} + \beta \chi_{it} + a_i + v_{it}$ and let us assume that there are only four periods that means $T = 4$ in this panel. Arellano and Bond in 1991 gave us the three moment conditions to estimate the model. As you know that this entire technology, entire methodology depends on method of moments.

So, if we assume $T = 4$ in this particular example, Arellano and Bond gave us three moment condition, what are those? $E[(y_{i,1}(v_{i,3} - v_{i,2})] = 0$. So, when you have $T = 4$ that means for third period you will have only $y_{i,3}$ and lag, when you go to fourth period then you will have both $y_{i,2}$ as well as $y_{i,1}$, $y_{i,2}$ will be following the Arellano and Bond's logic, $y_{i,t-1}$ means 4-2 is 2 $y_{i,2}, y_{i,1}$ is also available.

Because third order lag is also available in the fourth period that is why we will get moment condition with respect to $y_{i,1}$ and $y_{i,2}$, $y_{i,1}$ for both third and fourth period and $y_{i,2}$is only for the fourth period. So, accordingly three moment conditions. And then second-moment condition would be again and then $E[y_{i,1}(v_{i4} - v_{i3})] = 0$. And then third moment condition is $E[y_{i,1}(v_{i4} - v_{i2})] = 0$. So, these are the three moment conditions that we got from Arellano Bond.

(Refer Slide Time: 11:08)

Sargan, Hansen (J^{st},

$$E\left[\left(y_{i1}(v_{i3} - v_{i2})\right)\right] = 0$$
$$E\left[\left(y_{i1}(v_{i4} - v_{i3})\right)\right] = 0$$
$$E\left[\left(y_{i2}(v_{i4} - v_{i2})\right)\right] = 0$$

Any of these three moment conditions can be used to estimate $\hat{\rho}$. The remaining 2 moment conditions are overidentifying restrictions.

- Using Monte Carlo experiments, Bowsher (2002) found that the use of two many moment conditions cause Sargan test stat to be undersized and have extremely low power

And any of these three moment conditions can be used to estimate rho. As you know what is GMM basically? GMM is taking this population moment condition and converting into a corresponding sample moment condition and when you have more than one moment condition like the way we are getting here, the other remaining two moment conditions are over-identifying restrictions.

We need only one-moment condition, but we have three, so the remaining two are the overidentifying restrictions that we are putting in the model. Using Monte Carlo experiments, and since we have two moment conditions, we need to construct a test for overidentification and the first test is the Sargan test constructed within this particular context. Now the Sargan test what happens? Using the Monte Carlo experiments, how do you check this?

Either by the Sargan or by Hansen, sometimes this is known as J statistic, you have two tests when you have this type of overidentifying restriction to test. Now using Monte Carlo experiments Bowsher (2002) found that the use of too many moment conditions causes the Sargan test to be an undersized test, a statistic rather, a test statistic which follows a chi-square distribution actually, the Sargan test statistic to be undersized and have extremely low power.

(Refer Slide Time: 16:26)

Sargan test: (1) Reject the null (H_0) almost all the times when T is small

(2) Never reject the null (H_0) when T is large

Hansen test of overidentification:

Recall that the GMM Criterion fn or Quadratic loss fn:

$$Q = \left\{\frac{1}{N}\sum_i Z_i u_i(\beta)\right\}' W \left\{\frac{1}{N}\sum Z_i u_i(\beta)\right\}$$

What is the solution? What is the problem? In Sargan test what happens here when it becomes undersized and having extremely low power, then Sargan test what it does, Sargan test does two things. Reject the null almost all the time when T is small and secondly never rejecting the null when T is large that is the problem. This is the problem of the Sargan test and then when you go for a Hansen test of overidentification.

It depends on a function which we have discussed earlier, recall that the GMM criterion function or quadratic loss function that we have used earlier $Q = \left\{\frac{1}{N}\sum Z_i u_i(\beta)\right\}' W\{\frac{1}{N}\sum Z_i u_i(\beta)\}$, so this is the quadratic loss function. Hansen test depends on these quadratic loss function, how? This is the quadratic loss function.

If you go back, you can see this quadratic loss function, instead of u_i what we have defined earlier $y_i - \chi_i b$, here $y_i - \chi_i b'$ and W is a l by l weight matrix.

(Refer Slide Time: 20:55)

$$Q = \left\{\frac{1}{N}\sum_i z_i u_i(\beta)\right\}' W \left\{\frac{1}{N}\sum z_i u_i(\beta)\right\}$$

Then the test of overidentification is very simple.

if W is an optimal weight matrix, under the null hypothesis $H_0 : E\{z_i u_i(\beta)\}$,

the test statistic

$$J = N \times Q \sim \chi^2_{(e-k)}$$

for the test to be valid, W must be optimal $\Rightarrow$ W must be the inverse of the covariance matrix of the moment conditions

$$\Rightarrow \quad W^{-1} = E\{z_i u_i(\beta)\, u_i'(\beta) z_i'\}$$

Then the test of overidentification is very simple and how do you do this? If W is an optimal weight matrix under the null hypothesis which is H_0, what is the null hypothesis $H_0 H_0 = E\{Z_i u_i(\beta)\}$, the test statistic which is (J in the context of Hansen) $J = N \times Q \sim\chi^2_{e-k}$ here Q is basically a quadratic loss function.

For the test to be valid W must be optimal. That means that W must be the inverse of the covariance matrix of the moment condition. So that means basically what I am saying this implies $W^{-1} = E\{Z_i u_i(\beta) u'_i(\beta) Z'_i$, this is how the Hanse test of overidentification is defined and it depends on the calculation of the test statistic in Hansen test of overidentification basically depends on the quadratic loss function.

And what is the quadratic loss function, this is my quadratic loss function. **(Video Starts: 26:05)** And the advantage of this Hansen test in almost all the cases Hansen test is saying that overidentifying restrictions are valid. So, what we have learnt today then? We have learnt today how to use that *xtabond2* command to estimate a system GMM model and then we have also learnt how the quality of the estimate changes drastically if we change the model specification.

That means the difference GMM which was not ensuring the estimates to lie within the interval of FE and OLS estimates upper and lower bound, the moment we changed wage and capital also to be endogenous it changed drastically and pushed the estimate to lie within the interval. And then we learned about system GMM and we have also talked about the second post estimation checkup which is given after this *xtabond2* command which is by the Sargan and Hansen test.

And then we showed that the Sargan test it has some problems **(Video Ends: 27:40)** in a sense in many cases the Sargan test actually what we have written here look at here what it said that there are two tests to test for overidentifying restrictions are valid or not, one is Sargan, second one is Hansen and then this use of too many moment conditions that means when you have a relatively large time period.

That means if you have too many instruments when you are using, then the Sargan test is undersized and the power of the test also becomes extremely low. And that is why what it does? It rejects the null almost all the times when the T is small and almost never rejects the when T is large. That is the problem of the the Sargan test and then we go for the Hansen test of overidentification which is basically depending on the quadratic loss function.

This is the second post estimation checkup, but one thing then the most important thing what we have to keep in mind is that whatever model we estimate either system or difference GMM, the specification of the model should be correct in terms of what is exogenous, what is in an endogenous variable. So now there must be a test actually **(Video Starts: 29:11)** in Stata what we are doing here if we look at the command how we are writing.

We are specifying the model first and then we are specifying which is endogenous, which is exogenous by this GMM style and IV style. So, there must be a test for whether this specification is correct or not and that is also given in Stata if you look at the Sargan and Hansen tests they are also telling you the difference in the Hansen test of exogeneity, so we will talk about this thing later. With this, we are closing our discussion today.

And in our next class we will all be discussing some of this post estimation checkups and we will also learn how to control the number of lags because when too many instruments are used, we are using too many moment conditions because here we have no control, we have not specified on the number of controls, so look at here we are using how many controls from first period to eighth period of n, w and k. **(Video Ends: 30:37)**

And that is why the number of instruments becoming exponential and the number of instruments used can also determine the quality of estimates. So how to control for the number of lags that we learn later on. So, with this, we are closing our discussion today. Thank you.